Jesus Mancilla

Applied Scientist | AI/ML Systems Architect
LLMs, RAG, Evaluation, Vector Search, Retrieval Pipelines, Autonomous Agents, AI Systems Architecture
+1 650 391 4301 | jesus@jgmancilla.com | LinkedIn.com/in/jegama | GitHub.com/jegama | jgmancilla.com

Professional Summary

Applied Scientist building production-grade **LLM/ML systems** end-to-end: **vector databases**, **retrieval pipelines**, **RAG**, **evaluation/observability**, and **autonomous agents**. Blend of industry impact (Meta, Roku, Walmart, Argomai) and academic rigor (peer-reviewed publications). Track record translating business goals into scalable architectures with clear metrics: **cut multipage doc classification from ~90 min to <5 min** and **reduced weekly reporting from 6 hrs to <1 hr**; accelerated survey analysis from ~30 hrs to <8 hrs via hybrid ML.

Technical Stack & Domains

LLMs/ML: Retrieval, RAG, Evaluation, Prompting, Embeddings, Transformers, Scikit-learn, PyTorch, TensorFlow
NLP: Text classification, semantic search, vector stores, clustering, ranking
Data/Apps: Python, SQL, FastAPI, LangChain/LangGraph, React/Next.js, JS/TS
Systems: Enterprise architecture, service boundaries, domain models, NFRs, workflow automation
Cloud/Tooling: CI/CD, observability, Docker, Git
Other: Data viz, statistical inference; English & Spanish

Work Experience

Argomai — Houston, TX (Remote)
Senior Applied Scientist (Consultant) — Jan 2025 – Present

- Led enterprise **GenAI/ML architecture** (domains, service boundaries, data governance, NFRs) for customer-facing products; aligned initiatives with exec roadmaps.
- Engineered **vector databases**, **retrieval pipelines**, and **autonomous agents**; shipped a reusable component library (embeddings, prompt templates, orchestration SDK).
- Cut multi-page document classification from ~**90 min** → **<5 min**; reduced stakeholder reporting from **6 h/wk** → **<1 h** through automation.
- Applied abstraction & modularization to scale solutions across clients; improved latency/throughput consistency while lowering ops touchpoints.

Meta — Houston, TX (Remote)
Senior Quantitative UX Researcher — Jan 2024 – Jan 2025

- Built a **hybrid ML classifier** for open-ended responses (clustering, few-shot, human-in-the-loop, multi-agent reasoning) to triage/analyze surveys at scale.
- **Reduced analysis time from ~30 hrs to <8 hrs**; productized Python analytics tooling (stats, pipelines) for org-wide reuse.
- Fused behavioral editor logs with in-app surveys to create robust, multi-source datasets for model evaluation and product decisioning.
- Led longitudinal measurement for a new ML product with **500M+ MAU**; defined KPIs and bi-weekly tracking to inform iteration.

Roku — San Jose, CA
Senior User Experience Researcher — Jan 2021 – Nov 2023

- Developed the **Modular Survey Analysis System**: end-to-end **ML report generator** (stats + NLP categorization, clustering) over large-scale survey data.
- Created an **AI-powered indexed database** of UX & CI research; **cut report generation from ~4 hrs to <5 min** with ML automation and LLM summaries.
- Analyzed behavioral logs across **70M+ devices** to drive product decisions with quant evidence.

Walmart Global Tech — Sunnyvale, CA
Senior User Experience Researcher — Aug 2019 – Nov 2020

- Established KPIs and analytics for Sam's Club mobile; integrated interaction telemetry with business metrics to inform roadmap and experimentation.

Scrapworks Inc. — Palo Alto, CA
Data Scientist — Sep 2017 – Aug 2019

- **60% reduction** in forecasting error using deep learning for commodities futures; shipped dashboards over 20 years of sales data (**+30% sales growth**).
- Initiated NLP merchandise classification (supported a patent filing); productionized multi-source data ingestion/cleaning.

Additional Experience

Suggestic Mexico City, Mexico
User Experience Researcher Dec 2016 – Sep 2017
- Data-driven testing and analysis for new app features; rapid prototyping informed by analytics.

Stanford University Stanford, CA
User Experience Researcher May 2016 – Nov 2016
- Researched stress detection using ML over multimodal data (>150 hours of car/biometric/video); contributed to 90% accuracy algorithms.
- Co-authored papers on automotive UI and pedestrian interactions.

ITAM Mexico City, Mexico
User Experience Researcher Aug 2014 – May 2016
- Custom data visualizations and ML analysis on psychophysiological signals; personalized UX solutions across wearable, mobile, & web.

Selected Projects (Applied ML/LLMs)

Research Librarian (AI Index for UX/CI) — Semantic retrieval over an indexed research corpus using embeddings, vector stores, and custom ranking; improved findability and reuse across the org.

Modular Survey Analysis System — End-to-end ML pipeline that auto-generates survey reports (stats + NLP categorization + clustering); enabled at-scale analytics with minimal analyst time.

Customer Support Bot (Ref Architecture) — LLM-augmented support with RAG, evaluation/observability, and safety rails; blueprint for production conversational flows.

Education

Instituto Tecnológico Autónomo de México (ITAM)
M.S. in Computer Science (HCI/AI Focus) 2014 – 2016

Universidad de Colima
B.A. in Psychology 2009 – 2013

Selected Publications

Ramos-Rivera, R. E., Santana Mancilla, P. C., Garcia-Mancilla, J., & Gaytán-Lugo, L. S. (2025). Language models in education: Generative AI to optimize teacher performance analysis. InnovAcademica, 1(2), 74–85.

Ramos-Rivera, R. E., Garcia-Mancilla, J., Cárdenas-Villa, G. E., & Santana-Mancilla, P. C. (2024). Towards Improving Teacher Performance Assessment through Human-Centered AI-Powered Survey Analysis: An Approach Using Large Language Models (LLM). Avances en Interacción Humano-Computadora, 9(1), 261-264.

Baltodano, Sonia, Jesus Garcia-Mancilla, and Wendy Ju. "Eliciting Driver Stress Using Naturalistic Driving Scenarios on Real Roads." In Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, pp. 298-309. ACM, 2018.

Currano, Rebecca, So Yeon Park, Lawrence Domingo, Jesus Garcia-Mancilla, Pedro C. Santana-Mancilla, Victor M. Gonzalez, and Wendy Ju. "¡Vamos!: Observations of Pedestrian Interactions with Driverless Cars in Mexico." In Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, pp. 210-220. ACM, 2018.

J. Garcia-Mancilla, J. E. Ramirez-Marquez, C. Lipizzi, G. T. Vesonder, and V. M. Gonzalez, "Characterizing negative sentiments in at-risk populations via crowd computing: a computational social science approach," International Journal of Data Science and Analytics, Jun. 2018.

For full list, see: jgmancilla.com/research-papers